Journal of Transport Geography 72 (2018) 201–210

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Using points-of-interest data to estimate commuting patterns in central Shanghai, China

Mengya Li[a,b], Mei-Po Kwan[b,c], Fahui Wang[d], Jun Wang[a,*]

[a] Key Laboratory of Geographic Information Science (Ministry of Education), School of Geographic Sciences, East China Normal University, Shanghai 200241, China
[b] Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[c] Department of Human Geography and Spatial Planning, Utrecht University, 3508 TC Utrecht, The Netherlands
[d] Department of Geography and Anthropology, Louisiana State University, Baton Rouge, LA 70810, USA

ARTICLE INFO

Keywords:
Commuting pattern
Gravity model
POIs data
Central urban area
Shanghai

ABSTRACT

Commuting is an essential part of urban life. Long commutes have negative impacts on individuals and society, such as stress, loss of productivity, traffic congestion and air pollution. However, researchers often face the challenge of lack of data such as commute distance, duration, departure/arrival time, and origins/destinations in countries such as China. This study uses points of interest (POIs) to estimate employment locations, and implements a gravity-based model to estimate interzonal commuting patterns in central Shanghai, China. The results reveal a "busy corridor" in the west of the central city, especially during the morning peak hours. This pattern corresponds well with reported real-time traffic conditions in Shanghai. Our methodology offers a promising alternative for studying commuting patterns when such data are limited.

1. Introduction

Commuting is the movement of workers between residence and workplace. Intraurban patterns of commuting can be explained by the spatial separation between jobs and residences, and by the characteristics of commuters (Wang, 2000). Many studies have examined the interaction between commuting patterns and urban form (Cervero, 1989; Cervero and Wu, 1997; Giuliano and Small, 1993; Crane, 2000; Sultana, 2002; Sohn, 2005; Wang and Chai, 2009; Zhou et al., 2014). Some (e.g., Cervero, 1989) argue that jobs-housing balance significantly influences commuting, and better balance can relieve several pressing and persistent concerns in U.S. metropolises, such as congestion, energy depletion, air pollution, sprawl, and class segregation. Others (e.g., Giuliano and Small, 1993) contest that jobs-housing balance has minor or no effects on commuting, because individual workers' choice of residence and workplace depends on factors more important than job access, such as neighborhood quality, availability of green space, and racial and ethnic mix. Another notable strand of literature focuses on commuting behaviors in relation to commuters' socioeconomic attributes including gender (McLafferty and Preston, 1991; McLafferty, 1997; Kwan, 1999; Sermons and Koppelman, 2001; Kwan and Kotsev, 2015), occupation (Sang et al., 2011), income (Shen, 1998; Ibeas et al., 2013), marital status, and household type (Lee and McDonald, 2003; Sultana, 2005).

On data sources, there are three types for commuting studies. The first is survey data including official government censuses and informal surveys by researchers. The Census for Transportation Planning Package (CTPP) is a popular data source in the U.S. For example, Wang (2000) used the CTPP 1990 to explain the variations of commuting distance and time across traffic analysis zones (TAZ) in Chicago. Sang et al. (2011) used the CTPP 2000 to construct a disaggregate journey-to-work model for examining differences in average commuting distance by gender and occupation in Rochester, Minnesota. Habib et al. (2009) used the Transportation Tomorrow Survey 2001 to analyze commute schedules in the Greater Toronto Area. Merriman et al. (1995) employed the Japanese Population Census 1985 to measure excess commuting in the Tokyo Metropolitan Area. In China, for lack of such official datasets, Wang and Chai (2009) conducted face-to-face interviews with 736 households to examine the relationship between job-housing balance and commuting in Beijing, and Zhou et al. (2014) surveyed over 50,000 residents to analyze the commuting patterns in Xi'an. Both studies emphasized the role of *danwei* (a workplace-housing integrated unit inherited from the centrally-planned economy era in China) on commuting.

Secondly, researchers used various models to estimate data related to commuting trips. As commuting trips are seldom altered or canceled

* Corresponding author at: Key Laboratory of Geographic Information Science (Ministry of Education), School of Geographic Sciences, East China Normal University, 500 Dongchuan Road, Minhang District, Shanghai 200241, China.
E-mail addresses: fwang@lsu.edu (F. Wang), jwang@geo.ecnu.edu.cn (J. Wang).

https://doi.org/10.1016/j.jtrangeo.2018.09.004
Received 30 December 2017; Received in revised form 15 June 2018; Accepted 11 September 2018
0966-6923/

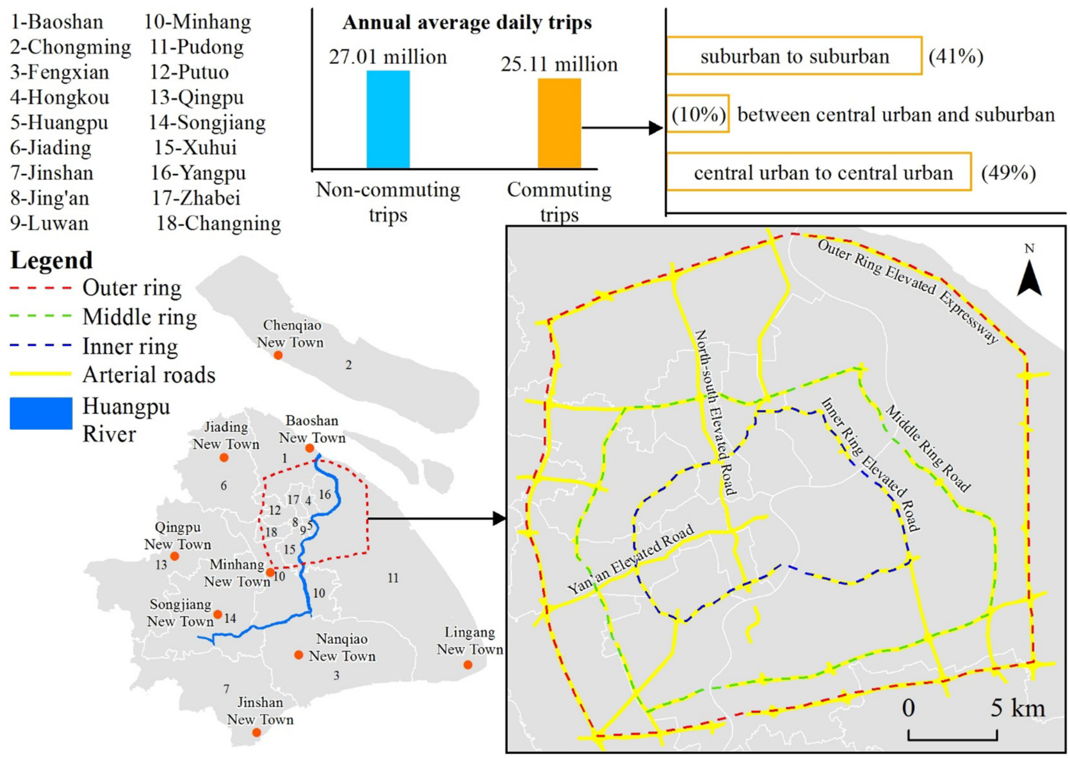

Fig. 1. Central urban area of Shanghai and the annual average daily commuting trips (2-column).

because of one's job obligation, commuters are more likely to encounter adverse events such as bad weather and high exposure to air pollutants (Kaur et al., 2007; De Nazelle et al., 2012; Li et al., 2015). However, this regularity is a prerequisite of modeling commuting flows produced by a large and complex worker group. Selection of models depends on the research questions and the geographic scale. The gravity model was a popular choice for measuring accessibility of stores, car parks, health cares, and jobs (Hansen, 1959; Luo and Wang, 2003), and was thus also widely used for estimating interzonal traffic (or trips distribution) at various scales. It emphasizes the distance decay effect on the mutual gravitation between two zones (e.g., less commuting between farther places and more commuting between populous places) (Sohn, 2005). The model requires readily available data on the origin and destination sizes (i.e., jobs and resident workers) as well as the travel distance or time matrix. Activity-based models, often implemented as structural equations, are developed for studies on a small scale to reveal individual behavior in relation to socioeconomic attributes. Kuppam and Pendyala (2001) used activity-based travel survey data and structural equations models to reveal the relationship between commuters' socio-demographic characteristics and their activity engagement and travel patterns in Washington DC. Cheng and Chen (2015) used a similar approach to study the travel behavior of low-income commuters in Nanjing.

The third group of studies relies on positioning techniques and big data. Big geospatial data afford us new opportunities for understanding our socioeconomic environments (Liu et al., 2015), including commuting patterns. For example, based on one-week smart-card data of bus rides in Beijing, Long et al. (2012) mapped the commuting flows between residential communities and business zones, and found the results to align well with the household travel survey data. Zhou et al. (2016) utilized mobile signaling data to analyze the impact of jobs-housing spatial mismatch on commuting behavior in Shanghai. Using a seven-day GPS dataset in Beijing, Shen et al. (2013) investigated the intra-personal day-to-day variability and flexibility of commuting behavior through 3D geovisualizations. To date, large-scale geolocation data or social media check-in data (e.g. Twitter, Facebook, Weibo) have been extensively used to predict urban mobility and travel behavior (Crampton et al., 2013; Hasan and Ukkusuri, 2014; Shelton et al., 2015; Poorthuis and Zook, 2017), but not yet used to distinguish commuting from travel activities in general. The main reason is for lack of identification of trip purposes in such data. Some attempts have been made to adopt multi-source data to overcome this limitation. For example, Bao et al. (2017) combined smart card data and Points-of-Interest (POIs) data to investigate the bike-sharing travel patterns in New York City. By analyzing surrounding POIs of a destination station, they were able to approximately classify trip purposes. Some remain suspicious of the reliability of knowledge derived from big data, and fear that it could be "an artifact of the algorithms used than the data itself" (Kwan, 2016: 275).

China does not collect official data like the CTPP in the U.S. on travel activities, and information on employment locations is also limited. The aforementioned informal survey or big data can be very costly and time consuming, and not feasible for most researchers. Data limitation is a major obstacle for commuting studies in China. This study sets out to offer an alternative data mining method for studying commuting patterns. Specifically, this paper uses points-of-interest (POIs) data, an open data source readily available to most researchers, to estimate employment locations and then models the commuting patterns. While our work is in Shanghai, the approach may be replicated for any cities including those second- or third-tier cities where paucity of commuting data is even more severe in China. Furthermore, most existing studies on commuting patterns in Shanghai and other large cities in China use large analysis unit such as district. Our work focuses on the central urban area (CUA) of Shanghai in a finer scale.

2. Study area

Over the past four decades, Shanghai has undergone significant spatial and social transformations during the urbanization process. By the end of 2016, there are over 24.2 million residents, circa 10 million of which are migrant workers who make a living in the city (http://www.stats-sh.gov.cn/tjnj/tjnj2017.htm). Residents between the age of